SUMMARIZATION DURING TUTORING: IMPLICATIONS FOR DEVELOPING MICRO-ADAPTIVE TUTORING SYSTEMS

Sandra Katz, Patricia Albacete, Pamela Jordan

Learning Research and Development Center, University of Pittsburgh

Paper presented at the 12th International Conference on Intelligent Tutoring Systems, Workshop 10, Pedagogy that Makes a Difference: Exploring Domain-Independent Principles Across Instructional Management Research within the ITS Community

Honolulu, HI, USA, June 2014

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130441 to the University of Pittsburgh. The opinions are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Summarization during Tutoring: Implications for Developing Micro-Adaptive Tutoring Systems

Sandra Katz, Patricia Albacete, and Pamela Jordan

Learning Research and Development Center, University of Pittsburgh, Pittsburgh PA, USA
{katz, palbacet, pjordan}@pitt.edu

Abstract. Although research on human tutoring highlights the importance of a high degree of *interactivity* between the tutor and student, some instructional strategies that could be carried out interactively are implemented didactically in tutoring systems. This is especially true of summarization, a ubiquitous instructional strategy. We investigated summarization during human tutoring in order to determine how to refine decision rules that specify when and how summarization takes place in a tutoring system so that these rules can be made more interactive and adaptive. We describe the informational requirements for carrying out these rules and implications for developing an authoring framework that can provide this information to a dialogue management system.

Keywords: human tutoring, natural-language tutoring systems, instructional strategies, summarization, authoring frameworks, GIFT, TuTalk, Rimac

1 Introduction

Several researchers have proposed that the large effect sizes of human tutoring can be attributed to its interactive nature—that is, the high degree to which the student and tutor respond to and build upon each other's dialogue moves [1], [2]. However, an important line of research conducted to test this *interaction hypothesis* shows that it is neither how much interaction takes place during automated tutoring that is important, nor the granularity of interaction. Instead, what matters most is how *well* the interaction is carried out—for example, what content the tutoring system addresses, when (e.g., in the context of which activities?) and how (e.g., using which tutoring strategies? delivered through which types of media?) [3, 4]. In other words, interactivity during tutoring needs to be carefully managed.

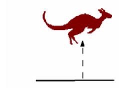
These important findings present several challenges to learning scientists and developers of natural-language tutoring systems. First, we need to determine which tutoring strategies are effective and for which types of learners. Second, because tutoring is essentially a linguistic process, we need to identify specific linguistic mechanisms that carry out these strategies. Several studies in the past decade have identified inctive dialogue patterns during tutoring whose frequency predicts positive learning outcomes (e.g., [5, 6]). Third, we need to specify decision rules that can guide a tutoring system in carrying out pedagogical strategies adaptively.

adfa, p. 10, 2011.

Table 1. Example of a post-problem reflective dialogue in Rimac

Problem: A kangaroo can jump about 2.5 m straight up. What is the magnitude of the take-off velocity?

Reflection Question: At what time during the jump (from the moment it leaves the ground to the moment it reaches the top of the jump) is the kangaroo moving fastest?



S = student; T = automated tutor (… = deleted dialogue)

S1: at the highest point *(incorrect response)*

Remedial dialogue implements Rule 3 in Table 2:

T2: Let's step through the reasoning. What is the magnitude of the kangaroo's velocity at take-off?

S3: .54m/s

T4: Good. Does the kangaroo's velocity change over the course of its upward jump or does the velocity stay the same?

S5: it gets smaller *(correct response)*

T6: what physics quantity represents change in velocity over time?

S7: acceleration *(correct response; triggers Rule 3)*

T8: Correct! Acceleration represents the change in velocity over time. In this case the acceleration is caused by gravity. This acceleration is pointing downward and it causes the kangaroo's upward velocity to get smaller…

End-of-tutoring summary

T9: To sum up, the kangaroo's velocity is greatest at take-off. It slows down as the kangaroo rises, until its velocity reaches 0 m/s at the top of the jump. Acceleration causes this change in velocity (velocity slows down to 0 m/s) and the acceleration is due to gravity.

Our research team has been addressing these issues in the process of developing Rimac, a natural-language tutoring system that scaffolds students in acquiring a deep understanding of physics concepts and principles, by engaging them in qualitative "reflective dialogues" after they solve problems and study worked examples [6]. (See Table 1.) Our goal is to specify tutoring decision rules that are empirically supported, domain independent, and more intuitive than those produced using automated approaches such as reinforcement learning (e.g., [3, 4]). Rimac's dialogues were developed using TuTalk, a dialogue development "toolkit" which has been used to build natural-language tutoring systems in various domains [7].

Our approach to deriving an initial set of decision rules to implement in Rimac can be summarized as follows. (See [6] for more detail.) We first identified patterns of collaborative dialogue exchanges in a large corpus of physics tutoring transcripts: 310 live tutoring sessions, in which one of seven tutors was paired with fifteen students. Interaction between the tutor and student was via teletype. We then conducted correlational analyses to identify relations whose frequency predicts positive learning outcomes and examined aptitude-treatment interactions. We described the context in which these potentially effective dialogue patterns typically occur and specified decision rules that capture relevant triggering conditions. We then implemented simplified versions of these rules within Rimac and are currently evaluating the system to determine whether these rules support learning—collectively, individually, and/or in groups defined by the tutorial strategy that they carry out. Table 2 shows a sample of these decision rules, expressed informally.

Table 2. Examples of tutoring decision rules to guide summarization in Rimac

End-of-tutoring "recap":

1. IF <*student response to problem or tutorial dialogue question* = correct>
   AND <*amount of steps or dialogue leading to response* = medium or high>
   ➔ Recap line of reasoning that led to correct response

Summaries during tutoring:

2. IF <*solution step or student response to tutorial dialogue question* = correct>
   AND <*amount of tutor scaffolding moves leading to response* = 0>
   ➔ Recap line of reasoning that led to correct response
3. IF <*current tutoring state* = remediation>
   AND <*student response to current tutorial dialogue question* = correct>
   ➔ Recap line of reasoning that follows from student response
4. IF<*student response to problem or tutorial dialogue question* = incorrect or partially correct>
   ➔ Summarize line of reasoning that leads to correct response
   OR scaffold student through main line of reasoning

These decision rules implement summarization—in particular, encapsulation of the line of reasoning that leads to a solution to a problem or answer to a question asked during tutoring, or that stems from a given problem-solving step. Our correlational analyses revealed that different types of line-of-reasoning summaries, as represented by these rules, predict learning [6]. For example, exchanges in which one dialogue partner (tutor or student) provides the steps in a line of reasoning that stem from, or lead to, a solution step expressed in his partner's turn predicted learning across ability levels (R=.65, p<.01), while tutor prompts for the student to summarize the reasoning that led to a correct answer to a problem or tutorial dialogue question predicted learning among high-knowledge students (R=.83, p<.05).

We take the view that language is purposeful action [8]. In essence, a summary abstracts the main points from a tutorial dialogue or any other instructional activity—for example, reading a text, watching a video, solving a problem, or playing an educational game. Summarization supports a wide range of instructional goals such as reinforcing facts and concepts, developing problem-solving scripts, and facilitating self-regulation of learning. It takes place dynamically, flexibly, and adaptively. As we illustrate presently, it can occur at the beginning of a tutoring session, at the end, and at various points in-between. It is typically didactic, but sometimes interactive.

In contrast, simulation of summarization in tutoring systems, including Rimac, is a far cry from capturing the level of flexibility and adaptability that we have observed during human tutoring. In response to this limitation and to the potential for summarization to support learning [9, 10], our goal is to make decision rules such as those shown in Table 2 more adaptive to students' cognitive and affective state. Towards this end, we are investigating summarization further in the tutoring literature and empirically. Although we also plan to do this with other tutoring strategies, this paper focuses on summarization—in particular, summarization of tutorial dialogue as opposed to other types of instructional activities. We first discuss what our analyses of summarization during naturalistic tutoring suggest about how summarization could be carried out more adaptively within tutoring systems. We outline the informational requirements of an adaptive tutorial dialogue system and describe how a domain-

neutral, interoperable ITS authoring framework such as the Generalized Intelligent Framework for Tutoring (GIFT, [11]) could be extended to support micro-adaptive implementation of summarization—that is, dynamic response to changes in students' cognitive and affective states.

2 Summarization During Human and Automated Tutoring

2.1 End-of-Tutoring "Recaps" of Main Points

Summarization happens in all forms of instruction—for example, end-of-chapter summaries in textbooks; recaps of classroom discussions or lectures, and of tutoring sessions with a human or automated tutor. In their extensive analyses of naturalistic tutoring sessions, Graesser and colleagues observed that summarization is one of the most frequent dialogue moves during Step 4 of their 5-step tutoring frame: scaffolding to improve an answer to a problem or question asked during tutoring [2], [12]. Frequent summarization is characteristic of both skilled and unskilled tutors [13, 14].

Natural-language tutoring systems, such as those developed within AutoTutor, typically simulate unskilled tutors' summarization practices, which Graesser et al. [14] describe as follows: "Unskilled tutors normally give a summary that recaps an answer to a question or solution to a problem. This summary serves the function of succinctly codifying a lengthy, multi-turn, collaborative exchange when a question is answered or a problem is solved" (p. 40). A tutoring system needs minimally adaptive decision rules to implement such end-of-tutoring recaps. The main parameter that determines whether a summary should be delivered is whether the topic under discussion has been adequately addressed during the dialogue, as reflected by the following AutoTutor decision rule [2]: "IF [quality of the cumulative collaborative exchange = completely correct] THEN [tutor supplies a summary or recap of the answer]" (p. 509).

Rule 1 in Table 2 similarly fires in Rimac when the student has arrived at a correct answer to a reflection question, after a series of questions that address the main ideas in the line of reasoning. See, for example, the summary at T9 in Table 1. Later versions of AutoTutor considered dialogue length, in addition to topic coverage, to determine whether to trigger a summary as the next dialogue move [12]: "IF [topic coverage = HIGH or number of turns = HIGH] THEN [select SUMMARY]" (p. 30). This rule ensures that the main points are extracted from lengthy dialogues and that short dialogues will not be summarized.

Typically, both skilled and unskilled tutors present end-of-tutoring summaries didactically. However, an alternative, which few tutors (skilled or unskilled) do, is prompt the student to generate a summary, perhaps with some degree of scaffolding from the tutor. Several tutoring system researchers have highlighted the potential benefits of doing so, in response to research which shows that generative activities such as summarization promote knowledge organization and retention [9,10]. To our knowledge, only one tutoring system, Guru, supports student summarization [13].

The choice between didactic and student-generated summarization provides a good example of how minimally adaptive decision rules such as those shown in this section

and in Table 2 could be refined to support more adaptive summarization during tutoring. Under what conditions should the tutoring system choose each option? One relevant factor is the student's level of knowledge of the topic discussed. For example, if the student made few errors, or the dialogue was short, a brief didactic summary (or none) would be appropriate. Another important factor is dialogue history [15]; in particular, prior exposure to a didactic summary. If the tutor has already summarized the material covered at the end of a previous dialogue, the student might be ready to generate a summary, perhaps with scaffolding. In addition to refining summarization decision rules to select who generates a summary, these rules could be extended to select suitable presentation media: text, static graphics, and/or video. The main parameter to consider is type of content: does text suffice, or is visualization necessary? If the latter, is the material static or dynamic? Learner preferences and "styles" (e.g., is the student a more visual or verbal learner?) could also be considered.

Our analyses of summarization during human tutoring sessions revealed other types of summaries besides the didactic, minimally adaptive end-of-tutoring recaps of main points that pretty much define summarization in tutoring systems. We therefore suggest the need to broaden the view of what can be summarized to include anything that the tutor expects to be in the "world of discourse" that he or she shares with the student because it is, or will be, relevant to the current tutoring session. This can include the content of a lecture, lab, or textbook section; a conversation during a previous tutoring session—not just what was discussed during the current session. We present a sample of the types of summaries that we identified in the next section, in order to provide a "case study" of what more dynamic, micro-adaptive implementation of tutoring strategies would entail.

2.2 Summaries at the Start of Dialogue and Various Points Along the Way

Line-of-reasoning summaries throughout dialogues. Whereas Rule 1 in Table 1 fires only at the end of a tutorial dialogue or problem, after a correct solution has been reached (e.g., T9 in Table 2), Rules 2-4 represent summarization that can take place at various times. Like Rule 1, these rules are minimally adaptive because they mainly respond to the correctness of the student's answer to the tutor's current question. As we will illustrate presently, a higher level of adaptivity could be reached by taking other factors into account. Due to space limitations, we focus on Rule 3.

Rule 3 captures situations in which the tutor addresses an incorrect answer to a question asked during tutoring through a remedial sub-dialogue. At some point during remediation, the student answers a question posed by the tutor correctly. The tutor then completes the line of reasoning that would lead from the student's answer to a correct answer to the original question that triggered the remedial dialogue, as illustrated in Table 1. Here the student answers the Reflection Question incorrectly (S1). The tutor launches a remedial dialogue at T2. When the student answers correctly at S7, the tutor completes the line of reasoning at T8. The purpose of this summary appears to be expediency; the tutor doesn't want to spend so long in a remedial dialogue that the student loses track of the original question (the Reflection Question). The tutor returns to this question in the end-of-tutoring summary at T9.

In the human tutorial dialogues that we analyzed, these "tutor completion" lines of reasoning varied in their degree of interactivity. Sometimes the tutor would prompt the student for a few more steps in the line of reasoning that followed from the student's correct answer, instead of delivering all remaining steps. Once again, factors such as the student's confidence level and dialogue history should be considered, in order to refine this rule. If the student has difficulty getting through the dialogue, then fewer opportunities for failure (i.e., incorrect responses to the tutor's questions) might prevent the student from giving up. Also, if the dialogue history log indicates that the student has engaged in remedial dialogues about the same content in the past, then perhaps a summary would be appropriate in the current dialogue.

Summaries at the start of dialogue (or early on). Sometimes tutors state the most salient features of a problem or other instructional activity, or the main concepts and principles to apply, for example: "You need to think of these problems as the same in the sense that when you are dealing with problems where we have several forces acting the first equation that should enter your mind is Newton's 2nd law." Alternatively or in addition, the tutor might provide a sketch of the main steps to be taken to solve a problem. Cade et al. [16] refer to such "game plan" summaries as "highlighting" and note that they are characteristic of expert tutoring. Tutors might label the type of problem being addressed and their key features, compare the current problem with previous problems, and/or outline the main steps in order to help students do what domain experts do: classify a problem early on and invoke relevant solution schema [17]. Like end-of-tutoring summaries, these early-session summaries draw upon the tutor's expectations of a shared "world of discourse" with the student. Breakdowns in these expectations need to be repaired, as illustrated presently.

Tutors tend to offer highlighting summaries when a student has had limited exposure to a given type of problem or displayed difficulty solving similar types of problems in the past. Consistent with these knowledge state attributes, highlighting summaries are typically delivered didactically, although the tutor might guide the student in co-constructing a "game plan" at the start of a problem-solving session after the student has demonstrated increased skill in solving similar problems, for example:

T: Let's start from the beginning. Use Newton's Second Law. What does this law say?

S: Fnet = mass * acceleration

T: That is the equation you need to use. Now what are the forces acting on the object?...

Another type of summary that is presented early on in a tutoring session is a "mini lecture" about the domain content associated with a problem or other type of learning activity. Again, these summaries respond to the student's knowledge state. For example, we observed that physics tutors summarized a topic targeted by the current problem when they incorrectly assumed that the student had sufficient exposure to that topic. In automated tutoring, degree of exposure to domain content could be determined before a student starts to work on a tutoring system, through a questionnaire or pretest. This data could be used to macro-adaptively prime the tutoring system to provide "mini-lectures" about topics that the student has had limited exposure to.

Summaries to support self-regulation. Students sometimes exhibit poor learning habits during tutoring. Several physics tutors that we observed responded with corrective advice that might promote self-regulation—for example, reminders to memo-

rize often-used equations; to express equations in terms of variables first and instantiate later. After the student solved the problem, tutors sometimes summarized this self-regulatory advice, for example: “OK that is what I got too but...next time, write the equations out symbolically first and then assign values before you do the actual calculation, so we can both follow what you are doing.”

3 Implications for Supporting Micro-Adaptive Summarization

The central aims of the preceding section were to show that summarization during human tutoring goes well beyond the didactic, end-of-tutoring “recaps” that are typically implemented in tutoring systems and to specify the types of information that would be needed by a dialogue system to simulate more flexible, dynamic and micro-adaptive summarization. This includes information about the learner’s knowledge state about domain content addressed during the dialogue; local and global dialogue history—for example, what topics have been covered during the current dialogue and during previous lessons? Has the content been summarized during a previous session and, if so, how (e.g., didactically or interactively; using which media?). Is the student ready to generate a summary, perhaps with scaffolding? Information about the student’s affective and metacognitive state is also important—for example, what types of self-regulatory advice should be recapped after a learning activity?

A modular, service-oriented, domain-independent framework such as the GIFT [11] will support management of this complex array of instructional information. The main “workhorses” are the Trainee, Pedagogical, and Domain Modules, which respectively model the learner’s cognitive and affective states, make decisions about what to teach and how to teach it (e.g., through which tutoring moves, using which strategies? etc.) and instantiate the preceding with domain content. As developers of the GIFT extend this framework to support micro-adaptive tutoring, the roles of these modules and inter-modular communication will need to be clarified—in particular, what types of information messages will each module send and provide? In addition, GIFT developers will need to consider how students’ responses and initiatives during tutorial dialogues can be used as input to the Trainee Module. As several tutoring researchers have noted, students’ dialogue contributions are one of the best resources for diagnosing the student’s knowledge about a topic [18]. Extensions to the GIFT such as these will greatly support the development of tutoring systems that interact with students as effectively as human tutors, perhaps more so.

Acknowledgement. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305A10063 and R305A130441 to the University of Pittsburgh. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute or the U.S. Department of Education.

References

1. Chi, M.T.H., Siler, S., Jeong, H., Yamauchi, T., Hausmann, R.G.: Learning from Human Tutoring. Cognitive Sci., 25, 471-533 (2001)
2. Graesser, A.C., Person, N., Magliano, J.: Collaborative Dialog Patterns in Naturalistic One-on-one Tutoring. Appl. Cognitive Psych., 9, 495-522 (1995)
3. Chi, M., VanLehn, K., Litman, D., Jordan, P.: An Evaluation of Pedagogical Tutorial Tactics for a Natural Language Tutoring System: A Reinforcement Learning Approach. Int. J. Artif. Intell. Ed., 21, 83-113 (2011)
4. Chi, M., Jordan, P., VanLehn, K.: When is Tutorial Dialogue More Effective Than Step-based Tutoring? In: Trausen-Matu, S., Boyer, K. (eds.) ITS 2014. LNCS. Springer, Heidelberg, 210-219
5. Forbes-Riley, K., Litman, D., Purandare, A., Rotaru, M., Tetreault, J.: Comparing Linguistic Features for Modeling Learning in Computer Dialogue Tutoring. In: Lane, H.C., Kalina, Y., Mostow, J., Pavlik, P. (eds.). AIED 2007. LNCS. IOS Press, Amsterdam (2007)
6. Katz, S., Albacete, P.: A Tutoring System That Simulates the Highly Interactive Nature of Human Tutoring. J. Educ. Psychol., 105(4), 1126-1141 (2013)
7. Jordan, P.W., Hall, B., Ringenberg, M., Cui, Y., Rosé, C.P.: Tools for Authoring a Dialogue Agent That Participates in Learning Studies. In Luckin, R., Koedinger, K.R., Greer, J.E. (eds.). AIED 2007. IOS Press, Amsterdam (2007)
8. Searle, J.R.: What is a Speech Act? Perspectives in the Philosophy of Language: A Concise Anthology. In Black, M. (ed.) Philosophy in America, 615-628 (1965)
9. King, A. Comparison of Self-Questioning, Summarizing, and Notetaking-Review as Strategies for Learning From Lectures. Am. Educ. Res. J., 29(2), 303-323 (1992)
10. Ley, K., Young, D.B. Instructional Principles for Self-Regulation. Educ. Technol. R&D, 49(2), 93-103 (2001)
11. Goldberg, B.S., Brawner, K.W., Sottilare, R., Tarr, R., Billings, D.R., Malone, N.: Use of Evidenced-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies. I/ITSEC. Vol. 2012. No. 1. National Training Systems Association (2012)
12. Person, N.K., Graesser, A.C., Kreuz, R.J., Pomeroy, V., TRG: Simulating Human Tutor Dialog Moves in AutoTutor. Int. J. Artif. Intell. Ed., 12, 23-39 (2001)
13. Olney, A.M., D'Mello, S., Person, N., Cade, W., Hays, P., Williams, C., Lehman, B., Graesser, A.: Guru: A Computer Tutor That Models Expert Human Tutors. In: Cerri, S.A., Clancey, W.J. (eds.) ITS 2012. LNCS. Springer, Heidelberg (2012)
14. Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., TRG: AutoTutor: A Simulation of a Human Tutor. J. Cogn. Syst. Res. 1, 35-51 (1999)
15. Moore, J.D.: What Makes Human Explanations Effective? In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society (1993)
16. Cade, W.L., Copeland, J.L., Person, N.K., & D'Mello, S.K. Dialogue Modes in Expert Tutoring. In: Schultz, R., Gentzoglanis, A. (eds.) ITS 2008. LNCS, pp. 470-479. Springer, Heidelberg (2008)
17. Larkin, J., McDermott, J., Simon, D.P., Simon, H.A. Expert and Novice Performance in Solving Physics Problems. Science, 208(4450), 1335-1342 (1980)
18. Person, N., Graesser, A.: Fourteen Facts About Human Tutoring: Food for Thought For ITS Developers. In: AI-ED 2003 Workshop Proceedings on Tutorial Dialogue Systems: With a View Toward the Classroom. 335-344 (2003)